DOCUMENT RESUME

ED 185 073 TM 800 059

AUTHOR Rivera, Charlene; Lombardo, Maria
TITLE The Validation of the English as a Second Language Assessment Battery (ESLAB).
PUB DATE [79]
NOTE 22p.; Figures 1 and 2 have been removed due to small print size. Tables 2 and 3 may be marginally legible due to small print size.

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Bilingual Students; *Criterion Referenced Tests; English; *English (Second Language); Junior High Schools; Language Dominance; Language Proficiency; Language Skills; *Language Tests; Second Language Learning; *Test Construction; Test Reliability; Test Validity
IDENTIFIERS *English as a Second Language Assessment Battery

ABSTRACT

The development of the English as a Second Language Assessment Battery (ESLAB) was based on needs indicated by achievement and placement statistics. The selection of the proficiency skills (listening, speaking, reading, and writing) was supported by changing views in language assessment and by legislation. Andrew D. Cohen's language model was adapted as a theoretical frame of reference. The battery was developed with a recognition of the interrelationship of native and second language skills. Receptive and expressive components of ESLAB included the Oral Screening Test, Oral Competency Test, Aural Comprehension Test, Dictation Exercise, Structural Competency Test, and Informal Reading Inventory. To ascertain ESLAB's statistical stability, 59 Hispanic inner-city bilingual students in grades 7 and 8 were tested. Each of the tests was subjected to item analysis, involving item difficulty and item discrimination. Face, content, predictive, and concurrent validity were established. Internal consistency and interrater correlations for the expressive subtests were also attained. Overall, both the receptive and expressive subtests of the ESLAB pilot edition proved to be valid and reliable measures of language proficiency. (Author/GDC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Charlene Rivera

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

The Validation of the English as a Second Language Assessment Battery (ESLAB)

Charlene Rivera
InterAmerica Research Associates
Arlington, VA

Maria Lombardo, Ed.D.*
Boston University Bilingual Training Ctr.
Boston, MA

ED185073

TM 800 059

Rationale and Theoretical Framework

The English as a Second Language Assessment Battery (ESLAB) was developed to meet the need for a valid and reliable criterion-referenced instrument to assess the English language proficiency skills of secondary bilingual students. A bilingual student is defined as one who lives in a two language environment, regardless of how well he/she speaks the languages (Zintz, 1975). The instrument is designed to assess both receptive (Listening/reading) and expressive (speaking/writing) language skills. It was intended to be used to place students according to pre-established entry level categories in appropriate levels for meaningful instruction in English as a Second Language (ESL) and reading.

Need for Language Proficiency Assessment

Several interrelated issues contributed to raising the current concerns for language proficiency assessment. Among them were achievement, the haphazard placement of minority students in educational programs, legislative developments, and the personal experience of working with bilingual students and teachers. In establishing the frame of reference for the ESLAB study, each aspect was considered.

Achievement. The educational achievement of the linguistic minority student was found to be "consistently below the achievement of the total national age population" (National Assessment," 1977, p. 5). Among the contributing factors to this situation were and continue to be language, sociocultural differences, amount of schooling, as well as societal attitudes toward the non-native English speaker. This educational situation went virtually unacknowledged until a government analysis of the school achievement of Mexican-American students in the Southwest demonstrated that 8.1 was the average years of schooling for students 14 years of age and older (U.S. Comm., "Unfinished Education," 1972). Parallel school achievement problems have been reported for other Hispanic, Asian, and Native American groups (U.S. Comm., "Puerto Ricans," 1976; Lau vs. Nichols, 1974; U.S. Comm. "Social Indicators," 1978). Statistics on the school drop-out or "push-out" rate of students from non-English speaking backgrounds also underscored the intensity of the need for re-assessment of the educational situation ("The Way We Go To School," 1970; U.S. Comm., "Excluded Student," 1972; Steiner, 1974).

*Lombardo and Rivera are co-authors of equal importance.

Placement. Inappropriate placement as a result of indiscriminate English assessment procedures involving intelligence tests also drew attention to the educational needs of the linguistic minority student. A classic example of this practice was documented in Mercer's (1971) Riverside, California, study of the labeling process.

In the Psychological Testing of American Minorities (1975), Samuda explored the ramifications of relying on intelligence and achievement test results in the case of minority students. He analyzed the testing controversy and pointed out that although tests have not proven to be a panacea, they will most probably not be abolished.

Fishman et. al. in the Guidelines for Testing Minority Group Children (1964) identified three major issues when using standardized tests with minority students:

1. Test results may not reliably differentiate among minority group scores;

2. Test results may not be predictive for the minority group student as they may be for the middle-class student;

3. Test results should be interpreted by a professional who is aware of the language and sociocultural background of the students. Samuda summarized Fishman's concern in stating that "good test usage depends upon conscientious, methodological, and critical examination of test scores" (p. 17).

De Avila and Havassy (1974) stress that low level performance may only be a reflection of minimal knowledge of the language, non-mastery of English reading skills and/or unfamiliarity with the culture of the test rather than an index of actual intelligence. For these reasons, language and cultural background should be considered before using standardized intelligence and achievement measures as indices of potential. The level of English language proficiency and the degree of familiarity with testing procedures should be taken into account before using any type of standardized English achievement test which may be used for placement purposes.

Legislative Developments. The official recognition of language minorities occurred in 1968 when Congress passed the Bilingual Education Act - Title VII as an amendment to the 1965 Elementary and Secondary Education Act. This landmark legislation stipulated that a program of instruction be designed to "teach children in English and to teach in (the native) language so they can progress effectively through school" (Bilingual Education: An Unmet Need," 1976, p. 11). In this definition, it is obvious that language was a unique component to be considered.

Language was also found to be important in the education of limited English-speaking students in the 1974 San Francisco court

case of Lau vs. Nichols. In complying with the court's opinion, the San Francisco Unified School District with a citizen's task force designed guidelines for school districts to follow in the case of the students whose "home language is other than English." Some months later Congress codified the decision as part of the Equal Educational Opportunity Act of 1974 (Teitelbaum and Hiller, 1977) and the Office of Civil Rights adopted guidelines which have come to be known as the Lau Remedies ("Task Force," 1975). They specify that students through language usage questionnaires be identified as:

A. Monolingual speaker of the language other than English
B. Predominantly speaks the language other than English
C. Bilingual
D. Predominantly speaks English
E. Monolingual speaker of English

Based on the general category in which a student falls, educational programs are then designed and matched to student needs.

While the Bilingual Education Act revisions of 1974 and the Lau Decision (1974), at the federal level, support the efforts of bilingual educators, Congress continues to press for results which will validate the government's efforts to support bilingual students in the educational process. Specifically, statistical data that documents the effects of bilingual education is sought. For this reason, in the extension of the Bilingual Education Act through the Education Amendments of 1978, there is an effort to clarify who is to be serviced through bilingual education. Students who are eligible for bilingual education, according to the new regulations, are no longer defined as being of "limited English-speaking ability" but rather to be of "limited English proficiency" (Education Amendments, 1978, p. 69). This alteration in definition refocuses the previous emphasis on proficiency apart from literacy. In other words, language proficiency according to the amendments includes all language skills, i.e., listening, speaking, reading, and writing.

Experience with Bilingual Teachers. Prior to legislative mandates, the experience of working with bilingual students strongly demonstrated the need to organize an approach to language competency assessment. While several governmental agencies (Foreign Service Institute, FSI; Central Intelligence Agency, CIA; and the Civil Service Commission, CSC) have developed a model for assessment of second language competence for adults learning foreign languages (Jones and Spolsky, 1975), a uniform procedure has yet to be developed for educators attempting to assess language competence of students in bilingual and/or other programs of instruction for language minority persons. The reality in the past has been that the individual classroom teacher who attempted to diagnose language skills has done so on an individual basis rather than as part of an organized methodological procedure. Additionally, as previously discussed, the interrelated issue of identifying adequate as well as valid test instruments have complicated the process. For although numerous formal and informal instruments to assess language competence exist, they are seldom comprehensive or organically integrated in design (Silverman, Noa, and Russell, 1977; Gutierrez and Rosenbach, 1975). Also, few have been found to be

designed for secondary students. This has created problems for the classroom teacher desiring to assess students' linguistic competencies. Not only must skills be catalogued, but appropriate tests must be found. As a result, the assessment procedure employed by a majority of secondary bilingual teachers has been random. It has lacked uniformity in the skills considered significant and in the instruments used to measure them. Finally, the process has been found to be time-consuming and tedious for the classroom teacher who needed an expedient method for assessing the strengths and weaknesses of bilingual students (Gonzales, 1979).

The Selection of Language Proficiency Skills

To date, the assessment of language proficiency of bilingual students has been difficult for two major reasons:

1. Inadequate identification of specific language proficiency skills necessary to determine the bilingual student's ability to perform in a monolingual or in a bilingual classroom (Cummins, 1979); Gonzales (1979) confirms this perspective in stating, "There is no general agreement among educators and/or linguists as to what constitutes either the particular functionalities of language which impinge significantly on school learning, the structures which situationally may affect it, or the specific skills which should be the minima at a given age or grade" (p. 13).

2. Lack of identification of proven, valid, and reliable standardized and/or criterion-referenced tests that measure linguistic competencies of secondary bilingual students in native languages and in English (Silverman, Noa, Russell, 1977).

In analyzing the skills of listening, speaking, reading, and writing, Loban (1963), Horowitz and Berkowitz (1967), Chastain (1976), Wilkinson (1971), Wilkinson and Stratta (1970), Moscovici (1967), and Higgins (1978) indicated that there are interrelationships among the four language skill areas. The results of the cited studies demonstrated that the receptive and expressive language areas require mastery of separate and somewhat different skills.

What was to be assessed was considered through a review of commonly used terms in language assessment: language dominance, language proficiency, and communicative competence. Language dominance was found generally to refer to oral competence. Language proficiency, on the other hand, was found to refer to both oracy and literacy skills. Communicative competence was found to refer to the social and cultural knowledge an individual is presumed to have enabling him/her to interpret oral and/or linguistic forms in a specific situation. These distinctions were made in an effort to identify the areas of language that should be considered in the development of the ESLAB.

Other considerations included the selection of a testing philosophy and test type. Since the purpose of the ESLAB was to assess language

skills in the four skill areas, both discrete point and integrative philosophies were found to be relevant. Applying this dual approach to language assessment, a person's ability to communicate orally or in writing in a particular situation, as well as the ability to manipulate the components of language in oral and written form, were taken into consideration. One other major decision in the development of the ESLAB was the one to adopt a criterion-referenced format. Although a language proficiency instrument could be either norm- or criterion-referenced, the purpose of the ESLAB seemed more in tune with a criterion-referenced instrumentation. This decision influenced both the design and the development of the test.

An adapted version of Cohen's (1975) model (Figure 1) was presented as the theoretical frame of reference. This model was adopted because it demonstrated the interrelationships among the language areas. As shown in the model, oracy and literacy skills are components of the receptive and expressive language areas. The model exemplifies the interrelationships among the language skills, the language components, the language varieties, and the language domains. As illustrated, although the language components are essential for the communicative process, they should not be thought of in isolation from the specific purpose of using language at any one time. The language components necessary for producing meaningful language include knowledge of the sound system (phonemes) for oral language and comprehension of the orthographical system (graphemes) for written language. Language use requires familiarity with vocabulary (lexicon), internalization of grammatical structures or rules of language usage (syntax), and the ability to attach meaning to referents (semantics). Most important is pragmatics or the ability to process sequences of linguistic components and relate them to the broader context of experience (Oller, 1970). The other aspects of language that are illustrated in Figure 1 are the language domains or context within which language can be described and the language variety or type of language that can be used in diverse geographical locations or within the same speech community or communities (Fishman, 1972).

It is seen that the language skills considered essential for communication are listening, speaking, reading, and writing. In essence, the figure demonstrates the variety of potential interrelationships among the skill areas and the language components.

Language Processing

The theoretical aspects of the language processing skills and their relationship to language proficiency assessment were also described through a model, Figure 2. It was developed in order to illustrate the major cognitive functions necessary to process language. Both cognitive and affective processes were considered. The ability to receive and manipulate language for communication purpose requires that an individual have the ability to attend, perceive, and remember phonological, lexical, and syntactical aspects of received language (input). Through the succesful process of synthesizing the input and normal motor control, expressive languege (output) becomes possible (McLaughlin, 1978). While

Figure 1 Interrelationship of the Language Areas

and

Figure 2 Processing of L_1 and L_2

have been removed because the print is too small to film legibly.

this analysis is somewhat simplified, for those concerned with language proficiency, knowledge of the aspects involved in processing language is important. This consideration becomes particularly necessary when a student being assessed for language proficiency seems to evidence problems related to processing rather than to language itself.

The affective forces of self-concept, motivation, and attitude also influence to what extent individuals activate their potential to manipulate language as a tool for communication. These factors are further documented in second language (L2) acquisition studies (Gardner and Lambert, 1971; Ervin-Tripp, 1973; Taylor, 1974), where it is demonstrated that these internal forces greatly interact to influence the acquisition of competence in the native language (L1) as well as in a second language (L2). With regard to the affective components, it is useful to keep in mind that self-concept, motivation, and attitude can be modified through internal and/or external changes.

The ESLAB Pilot Test

The ESLAB instrument was developed with a recognition of the interrelationship of native and second language skills. Cummins (1979) provided understanding of this relationship through his theory of interdependence. He pointed out that "development of competence in a second language (L2) is partially a function of the type of competence already developed in L1, when intensive exposure to L2 begins" (Cummins, 1979, p. 222). Through several studies it was found that the abstract control of language was most important in facilitating second language acquisition. Literacy skills in the native langauge were found to play a particularly important role in arriving at a threshold level of language that would allow the potentially beneficial aspects of bilingualism to be exhibited. For this reason, although the ESLAB was designed to assess English language skills, it has been strongly recommended that its administration be accompanied by a native language examination.

The completion of the ESLAB pilot test version took approximately two years. The process involved the development of objectives and selection of a technique most appropriate to meet the stated objectives. The content base was selected from curriculum being used in ESL and bilingual programs. Linguistic and cultural backgrounds of the potential students to be assessed were also taken into account.

The materials developed for the pilot test version were the ESLAB Manual, Handbook, and Scoring Charts. Each component was briefly described. The ESLAB receptive components included the Aural Comprehension Test, the Structural Competency Test and the Informal Reading Inventory. The ESLAB expressive components included the Oral Screening Test, the Oral Competency Test, the Dictation Exercise, and the Writing Sample.

ESLAB Scores: Derivation and Interpretation.

In scoring the ESLAB the test administrator first tabulates student results. Once the student's performance is ranked according to the individual scoring criteria, the total points are compared to the

criterion for establishing ESL and reading levels (Table 1). Test components are all translated into levels except for the Oral Screening Test which is scored as pass/fail. Based on overall results, a language profile for each student is established for placing students into instructional groupings.

The rationale supporting the need for a hierarchically arranged entry-level criteria was based on curricular considerations, experience with bilingual students, the research which suggests that there is a threshold level of language competence which must be attained in order to avoid cognitive deficits and to allow the potentially beneficial aspects of bilingualism to affect cognitive functioning (Cummins, 1977; 1979), and the research indicating that there is a hierarchy of language skills (Chastain, 1976).

The levels are defined in terms of the test authors' experience in working with bilingual students. They are: Beginner I, Beginner II, Intermediate I, Intermediate II, and Advanced. Table 1 provides a summary of the criterion for the entry level categories.

Table 1
Criterion for Establishing ESL and Reading Levels

Level	ESL (L2)	English Reading
Beginner I	Illiterate student in L1 may or may not have aural/ oral L2 skills	-
Beginner II	Literate student in L1 may or may not have aural/ oral L2 skills	-
Intermediate I	Literate student in L1 has basic aural/oral English skills, limited L2 literacy skills	Reads at 2nd gr. level
Intermediate II	Literate student in L1 has basic aural/oral L2 skills, limited English literacy skills	Reads at 3rd-5th gr. level
Advanced	Literate student in L1, literate in English	Reads on grade level

The beginners levels are subdivided in order to accommodate two very differently equipped types of students. Beginner I is for students who may or may not be conversant in English. The common denominator for this

group is the lack of literacy skills in either the native (L1) or in the second language (L2). Beginner II includes students who may or may not possess aural/oral skills in English, but who are literate to at least a second grade level in their native language. Intermediate I students are defined as those students who have aural/oral skills in English and who capable of reading English at 2nd grade level. Intermediate students possess aural/oral English skills and are able to read in English at 3rd to 5th grade level. Advanced students possess aural/oral English skills and can read on grade level.

Description of the ESLAB Tests

Oral Screening (O.S.) Test. The Oral Screening Test is composed of 10 informational questions. In determining a student's performance, comprehension and production are evaluated on a parallel rating scale by the test administrator. Administered individually, it is tape recorded in order to facilitate scoring. Judgement is made on how well the student comprehends information and is able to respond appropriately. Comprehension is determined by a student's ability to respond accurately (but not necessarily in extended discourse). A parallel procedure is followed for evaluating a student's overall production. The overall performance level is based on the actual number of correct responses defined by a predetermined criteria. Scored on a pass/fail basis, it is recommended that students who pass the Oral Interview continue to the Oral Competency Test.

Oral Competency (O.C.) Test. The Oral Competency Test was based on a quasi-realistic activity and was designed to assess a student's degree of oral competency through the oral interpretation of a visual stimulus. Questions concerning the visual stimulus were developed by the authors in order to assess overall communicative competence. The test is administered individually and a tape recorder is recommended to facilitate evaluation of the student's discourse. The performance level is based on ratings of the student's overall control of phonology, vocabulary, grammar, and fluency as defined on a predetermined criteria. Upon completion of the test, all students, regardless of the score, continue with the Aural Comprehension Test. This is recommended because often in the case of second language learners aural comprehension far exceeds productive abilities (Chastain, 1976).

Aural Comprehension (A.C.) Test. The Aural Comprehension Test was specifically designed to assess listening comprehension. It was modeled on the Pillsbury, Thrasher, Upshur (1963) Aural Comprhension Test. The sentence structures in this test were selected as a guide because they were representative of those necessary to function in English. The ESLAB Aural Comprehension Test consists of 25 statements. Each statement is accompanied by four picture frams from which the student selects the appropriate picture. It is administered as a group test and scored according to the number of correct responses. The overall performance level is based on the actual number of correct responses as defined by the predetermined criteria. Regardless of the achieved level, it is recommended that students progress to the Dictation Exercise so that basic literacy skills level can be assessed.

Dictation Exercise (D.E.). Dictation involves the examination of a

student's English listening and writing skills or more specifically, a student's ability to reproduce sounds, structures, vocabulary, and stories. It consists of 12 words and a short paragraph. The words used were taken from the Dale-Chall (1948) word list. Administered as a group test, it requires that students write the list of words and a short paragraph. The performance level is determined by rating the student's control of word usage, spelling, and grammar as defined on a predetermined criteria. Oller's (1970) study indicated that the student's performance was an indication of the ability to recognize phonological and orthographical structure of the English language. Thus, if the student does not perform well on the dictation exercise, two possibilities should be explored: either the student is literate in his/her native language but needs aural/oral drill before being able to read and write in English, or the student is assumed to be illiterate in any language while he/she may or may not possess a high degree of aural/oral skills in English. Students functioning at a Beginner II level on both the Aural Comprehension Test and the Dictation Exercise are considered to be sufficiently literate to continue with the Structural Competency Test.

Structural Competency (S.C.) Test. The Structural Competency Test was developed to measure knowledge of English syntactical structures requiring students to visually identify the syntactical patterns of English. This test is modeled after the Michigan Aural Examination (Pillsbury et. al., 1963) and the English Sentence Structure curriculum developed by Krohn et. al. (1971). These references used by bilingual and ESL teachers as a guide for teaching English structures provided the guidelines for the structures selected. In its administration, the student is given a structure and four multiple choice responses from which to select. The student's performance is evaluated according to the actual number of correct responses as defined by a predetermined criteria. It is recommended that students functioning at an Intermediate I level or above continue to the Informal Reading Inventory (IRI).

Informal Reading Inventory (IRI). The ESLAB Informal Reading Inventory (IRI) consists of a set of eight original passages which range in length from 30 to 200 words. The Fry (1968) and Dale-Chall (1948) readability formulae were used to level the passages from primer to eighth grade. Ten multiple choice questions for each passage were constructed in order to evaluate the student's comprehension level. They were based on Barrett's Taxonomy (Smith and Barrett, 1974). The IRI is administered as a group test. Students read the graded stories silently and then answer the questions at the end of each selection. They are advised to stop when they no longer comprehend.

The overall reading level is then established using the following criteria: Independent: 95-100%; Instructional: 75-94%; Frustration: 0-74%. Students with an instructional level of at least second grade, Intermediate I level, are asked to complete the Writing Sample. This procedure is recommended because students who are not capable of reading should not be forced to independently construct a written passage.

Writing Sample (W.S.). Constructed as an unfinished story, The ESLAB Writing Sample starter was taken from the book The Spider Plant (Speevack, 1968). The test is administered as a group test and the performance

level is determined through an evaluation of the student's control of word usage, spelling, grammar, and organization.

Although the skills examined by the ESLAB tend to overlap, the tests have been arranged according to the difficulty involved in producing language. Students are asked to listen and speak, and then read and write. The ESLAB Tables of Specifications (Tables 2 and 3) outline the skills assessed in both the receptive and expressive areas.

Table 2
ESLAB Table of Specifications
Expressive Area

Skill	Test	Behavioral Objectives	Number of items per objective	Stimulus Mode	Response Mode	Test Type
Listening/ Speaking	Oral Screening	Given 10 oral questions, the student will be able to comprehend and verbalize appropriate responses.	10	Oral	Oral	Verbal Essay
	Oral Competency	Presented with a picture, the student will be able to respond orally to questions, indicating ability to comprehend and verbalize utilizing appropriate phonology, grammar, vocabulary, and fluency.	13	Visual/ Oral	Oral	Verbal Essay
Listening/ Writing	Dictation Exercise	Dictated a list of words, the student will be able to reproduce them correctly. Dictated a paragraph, the student will be able to remember phonemes, syntax and be able to reproduce these accurately in terms of word usage, spelling, and grammar.	30	Oral	Written	Reproduction
Writing	Writing Sample	Presented with an incomplete story, the student will be able to read silently, comprehend, and then write an ending to the story demonstrating understanding of word usage, spelling, grammar, and organization.		Written	Written	Written Essay

BEST COPY AVAILABLE

Table 3
ESLAB Table of Specifications
Receptive Area

Skill	Test	Behavioral Objectives	Number of items per objective	Stimulus Mode	Response Mode	Test Type
Listening	Aural Comprehension	Given 25 oral statements, the student will be able to speak the appropriate response from among four alternatives representing each statement.	25	Visual/ Oral	Marking	Multiple Choice/ Matching
Reading	Structural Competency	Given 50 incomplete sentences the student will be able to read silently and select the correct response among four grammatical alternatives representing each statement.	50	Written	Marking	Multiple Choice Completion
	Informal Reading Inventory	Presented with eight stories and 10 multiple choice questions (per story) the student will be able to encode, decode, and comprehend at levels which will indicate Frustration, Instructional, and Independent Reading levels.	80	Written	Marking	Multiple Choice

Validation of the ESLAB

In order to validate the ESLAB tests, pilot testing was conducted in an inner city area comprised of a large Hispanic population. Initially, the examiners consisting of bilingual and ESL teachers were trained in the administration and scoring of the battery, through three teacher training workshops. Then a sample of 59 Hispanic bilingual students ages 12 to 15 --seventh and eighth graders-- was tested. When all the data had been collected, the results for each of the seven tests were analyzed. Item analysis was performed and validity and reliability were established.

Item Analysis. Testing experts (Thorndike and Hagen, 1969; Popham, 1971; Litchman, 1973; Sax, 1974) varied in opinion as to the appropriate procedure for item analysis of criterion-referenced measures. For this reason, item analysis for the Receptive and Expressive areas of the ESLAB was based upon the test developers' judgement of appropriate procedures and empirical data. For the receptive area, the difficulty index (i.e., the percentage (p) of students correctly responding to an item) and the discrimination index (i.e., correlation between item and overall test results or the point biserial correlation (RPB)) were obtained. The p and the RPB values are described in Table 4.

Table 4
Item Analysis Summary - ESLAB Receptive Subtests

	p[a]	RPB[b]
Aural Comprehension Test	27.1-89.9%	.08-.67
Structural Competency Test	1.7-64.4%	-.01-.47
Informal Reading Inventory	0.0-69.5%	.00-.70

n = 59

[a]Percentage or difficulty index
[b]Point biserial correlation

The results indicated that items on the Aural Comprehension Test were easy in contrast to the results of the Structural Competency and Informal Reading Inventory. Overall student performance on the Aural Comprehension Test was high, indicating that students had mastered the content. The significantly lower performance on the other two tests indicated that students' aural skills were much stronger than was their ability to read and write.

For the expressive subtests which are evaluated on a rating scale, the analysis for each item was viewed according to how each student performed on the basis of the set criteria.

For the Expressive Area, item quality was indicated by a high correlation between subtests and the total score of each test (Nelson, 1974). Students were rated on a scale from poor to excellent. The results indicated whether the high achievers on each item were the high achievers on the specific test. The only test exempted from this process was the Oral Screening Test, which was rated on a pass/fail basis. The item analysis results are summarized on Table 5.

Table 5
Item Analysis Summary - ESLAB Expressive Subtests

	Ratings			
	Poor	Fair	Good	Excellent
Oral Competency Test	7.70	17.48	53.30	23.33
Dictation Exercise	14.70	25.80	35.03	6.62
Writing Sample	6.45	48.72	16.10	6.35

$n = 59$

Item discrimination for each of the four tests was significant at $p < .01$. Overall, students performed better on aural/oral tests than the literacy tests.

Validity. Validity is the extent to which a test measures what it purports to measure (Heaton, 1975). For the ESLAB, validity was based on curricular and empirical analyses (Green, 1975). Curricular considerations included face and content validity analyses. These were determined by having examiners and examinees evaluate the tests and by having language and reading experts examine each of the items. In most cases, items were rated favorably; where they were not, suggestions were made for altering the items. Empirical analysis was also of two types. The first was predictive validity for which Kendall's tau (Nie, et. al., 1975) was used to correlate each test component's level results (Tables 6, 7, and 8) with four Teacher Estimates (T.E.) and ESL report card grades. Results indicated positive correlations at $p < .01$ in most cases. The second type was concurrent validity (Table 9), used only for the IRI. A Pearson Coefficient indicated a high correlation or close relationship between the IRI and the other three reading tests indicating that they may have produced similar results. The Teacher Estimate had a negative correlation with the IRI possibly because teachers tended to underestimate students' reading levels.

Table 6

ESLAB Tests
Predictive Validity Correlations Between Teacher Estimates, Level Performance in the Expressive Language Area Tests, and ESL Grades

Teacher Estimates Teacher/Component		Level Performance O.C.	D.E.	W.S.	ESL◦ Grades
1	O.C.		.30893**		.41657**
	D.E.		.38769*		
	W.S.		.24947*		
2	O.C.	.34174**			
	D.E.	.44880**	.48658**	.32423**	.37693**
	W.S.	.37337**			.36109**
3	O.C.	.25211**			.32598**
	D.E.	.23859**			
	W.S.				
4	O.C.	.75003**		.45560**	
	D.E.			.51110**	
	W.S.				

Table 7

ESLAB Tests
Predictive Validity Correlations Between Teacher Estimates, Level Performance in the Receptive Language Area Tests, and ESL Grades

Teacher Estimates Teacher/Component		Level Performance A.C.	S.C.	IRI	ESL◦ Grades
1	A.C.			.25899*	.41850**
	S.C.		.33405**	.25874*	
	IRI				.35967**
2	A.C.		.35714*	.28360*	
	S.C.		.48658**	.32880**	.37693**
	IRI		.57050**	.34568**	.24301*
3	A.C.				
	S.C.		.32931**		.29860**
	IRI		.34397**		
4	A.C.		.65049**	.50581**	
	S.C.				
	IRI				

n = 59
*p<.05
**p<.01

◦English as a Second Language

Table 8
Predictive Validity Correlations: Receptive and Expressive ESLAB Subtests Results with ESL Grades

Receptive	ESL Grades	Expressive	ESL° Grades
A.C.	.37662**	O.C.	.25923**
S.C.	.25200**	D.E.	.29942**
IRI	.37600**	W.S.	.23347**

n = 59
**p<.01
°English as a Second Language

Table 9
Concurrent Validity for the ESLAB Informal Reading Inventory

	Cloze Test	Stanford Diagnostic Reading Test	Teacher Estimate
Informal Reading Inventory	.7484**	.6203**	-.0456
Cloze Test		.6105*	.3577
Stanford Diagnostic Reading Test			.4248*

n = 59
*p<.05
**p<.01

Reliability. Reliability is defined as the accuracy with which a test measures whatever it is intended to measure (Thorndike and Hagen, 1969). For the ESLAB, two types of reliability indices, internal consistency and interrater reliability were obtained. The index of internal consistency "is an index of the consistency of the subtests, or the degree to which the subtests measure the same thing" (Nelson, 1974, p. 280). In other words, internal consistency indexes whether a total test score measures common skills rather than individual distinct ones. In this study this means that all items should be measuring linguistic skills and not, for example, intellectual functioning, or attitudes. The index of reliability was determined for both the Receptive and Expressive Language Areas (Table 10) through computation of Hoyt's anova and Cronbach's alpha.

Table 10

Reliability for the Receptive and
ESLAB Expressive Subtests

Test				Coefficient Range				
Receptive	Items	Mean	S.D.[a]	Low	High	Hoyt's anova	SEm[b]	Cronbach alpha
Aural Comprehension	27	19.46	4.67	0	26	.81	1.97	
Structural Competency	52	9.92	3.04	0	15	.37	2.39	
Informal Reading Inventory	80	11.14	6.61	0	22	.83	2.74	.74
Total Test Statistics	159	39.78	9.31	18	58	.79	4.25	.21
Expressive								
Oral Screening	20	53.80	10.11	0	67	.92	2.75	
Oral Competency	4	11.76	2.39	0	16	.78	.96	
Dictation	3	7.73	2.92	0	12	.90	.74	
Writing Sample	4	7.08	4.22	0	15	.94	.86	
Total Test Statistics	31	80.37	13.89	19	99	.91	4.17	.67

$n = 59$

[a]Standard Deviation

[b]Standard Error of Measurement

Interrater reliability was established only for the expressive subtests. For this purpose, the reliability of the ratings of three raters were correlated on a sample of 30 students. This procedure was followed in order to ensure their reliability in assessing students. Approximately half of the students tested were randomly selected to be subjects for obtaining interrater reliability (Table 11).

Table 11
ESLAB Interrater Reliability Summary

	Mean	Standard Deviation	Standard Error of Measurement	Hoyt's Anova
Oral Screening	164.40	13.50	3.96	.91
Oral Competency	75.63	5.85	2.85	.76
Dictation	23.03	7.25	1.97	.92
Writing	26.97	3.50	2.34	.51

n = 30

On the whole, the interrater reliability coefficients were quite significant demonstrating a high degree of agreement among raters.

Overall, the seven subtests were reliable with the exception of the Structural Competency Test. On the basis of positive content and predictive validity results, it was retained as part of the entire battery with a strong recommendation for field testing.

Standard Error of Measurement (SEm). The SEm values for the Receptive Tests were Aural Comprehension 1.97, Structural Competency 2.39, and IRI 2.74 with an overall value of 4.25.

The Expressive Area results were: Oral Screening Test 2.75, Oral Competency Test .96, Dictation Exercise .74, and Writing Sample .86, with a total score of 4.17. In both instances, values were fairly low indicating the tests to be a fair assessment of student ability.

Significance of the Study

The importance of the study is that: (a) it meets the critical need for an integrated criterion-referenced language assessment battery; (b) it provides a basis for evaluating language areas that are significant for learning to read in English; (c) it identifies independent, instructional, and frustration levels for the purpose of grouping students in reading. Additionally, the use of the ESLAB provides curriculum developers with an understanding of achieved and non-achieved student skills, and school systems with potential evaluative information of their bilingual program.

References

Bilingual Education: An Unmet Need. Report to Congress. Washington, D.C.: U.S. Office of Education, 1976.

Chastain, Kenneth. Developing Second Language Skills. Chicago: Rand McNally Publishing Company, 1976.

Cohen, Andrew D. A Sociolinguistic Approach to Bilingual Education. Rowley, Massachusetts: Newbury House Publishers, 1975.

Cummins, James. Linguistic Interdependence and the Educational Development of Bilingual Children. Review of Educational Research, 1979, 49, (2), pp. 222-251.

Cummins, James. Psycholinguistic Evidence. Bilingual Education: Current Perspectives. Arlington, Virginia: Center for Applied Linguistics, 1977.

Dale, Edgar and Chall, Jeanne. A Formula for Predicting Readability. Educational Research Bulletin, January, 1948, 28, 11-20.

DeAvila, E. and Havassy, B. Some Critical Notes on Using IQ Tests for Minority Children. Materiales en Marchas. Texas: Dissemination and Assessment Center for Bilingual Education, 1974.

Education Amendments of 1978, A Report to accompany Senate Bill 1753. Senate Calendar 787, Report 95-856 95th Congress, 2nd Session, May 15, 1978.

Ervin-Tripp, Susan and Osgood, Charles E. Second Language Learning and Bilingualism. In Susan Ervin-Tripp (Ed.) Language Acquisition and Communicative Choice, Stanford, California: Stanford University Press, 1973.

Fishman, Joshua. The Sociology of Language. Rowley, Massachusetts: Newbury House Publishers, 1972.

Fishman, Joshua A., et al. Guidelines for Testing Minority Group Children. Journal of Social Issues Supplement, 1964, 20, 129-145.

Fry, Edward. A Readability Formula that Saves Time. Journal of Reading, 1968, 11, (7) 513-516.

Gardner, Robert and Lambert, Wallace. Attitudes and Motivation in Second Language Learning. Rowley, Massachusetts: Newbury House Publishers, 1971.

Gonzalez, Josue. Josue Gonzalez opens NABE Convention. NABE NEWS, 1979, 2 (5).

Green, J. A. Teacher-Made Tests. (2nd ed.) New York: Harper and Row, 1975.

Gutierrez, Medardo and Rosenbach, John. Bilingual Assessment Test Development and Reviews: A Manual for Teacher Use. New York: The State Department, 1975.

Heaton, J.B. Writing English Language Tests. London: Longman Group Limited, 1975.

Higgins, Tony E. Written Communication as Functional Literacy: A Developmental Comparison of Oral and Written Communication. In Richard Beach and P. David Pearson (Eds.) Perspectives on Literacy. College of Education, University of Minnesota, 1978.

Horowitz, Milton and Berkowitz, Alan. Listening and Reading, Speaking and Writing: An Experimental Investigation of Differential Acquisition and Reproduction of Memory. In Perceptual and Motor Skills, Southern University Press, 1967, 21, 207-215.

Jones, Randall L. and Spolsky, Bernard. Testing Language Proficiency. Arlington, Virginia: Center for Applied Linguistics, 1975.

Krohn, Robert, et al. English Sentence Structure. Ann Arbor, Michigan: The University of Michigan Press, 1971.

Lau vs. Nichols, 414 U.S. 563, 566 (1974). The U.S. Law Week Supreme Court Opinions. Washington, D.C. The Bureau of National Affairs, January, 1974, 42, 4165-4168.

Lichtman, Marilyn. The Development and Validation of R/EAL, an Instrument to Assess Functional Literacy. (ERIC Report No. ED 081 811), Washington, D.C.: Catholic University of America, June, 1973.

Loban, W.D. The Language of Elementary School Children Research Report No. 1. Champaign, Illinois: National Council of Teachers of English, 1963.

Lombardo, Maria. The Construction and Validation of the Listening and Reading Components of the English as a Second Language Assessment Battery. Boston University, unpublished doctoral dissertation, 1979.

Madden, R., Karlsen, B., & Gardner, E.F. Stanford Diagnostic Reading Test. New York: Harcourt Brace, Jovanovich, 1976.

McLaughlin, Barry. Second Language Acquisition in Childhood. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc., 1978.

Mercer, Jane. The Labeling Process. Paper Presented at John F. Kennedy Center for the Performing Arts, Washington, D.C., October 16, 1971.

Mascovici, S. Communication Processes and the Properties of Language. Advances in Experimental Social Psychology, 1967, 3, 225-270.

National Assessment of Educational Progress. Hispanic Student Achievement in Five Learning Areas: 1971-75. Washington, D.C.: U.S. Government Printing Office, May, 1977.

Nelson, L. R. LERTAP. Boulder, Colordao: Laboratory of Educational Research, University of Colorado, 1974.

Nie, N.H.; Hull, C.H.; Jenkins, J.G.; Steinbrenner, K.; Bent, D.H. New York: McGraw-Hill, 1975.

Oller, John W., Jr. Linguistics and the Pragmatics of Communication, 1970. (ERIC Document Reproduction Service No. ED 041 292.)

Pillsbury, Paul W., Thrasher, Randolph, and Upshur, John. Aural Comprehension Examination: ELI English Achievement Series. Ann Arbor, Michigan: The University of Michigan, 1963.

Popham, W. James. Indices of Adequacy for Criterion-Referenced Test Items. In W. James Popham (Ed.) Criterion-Referenced Measurement. New Jersey Educational Technology Publications, 1971, 79-98.

Rivera, Charlene. The Construction and Validation of the Speaking and Writing Components of the English as a Second Language Assessment Battery. Boston University, unpublished doctoral dissertation, 1979.

Samuda, Ronald J. Psychological Testing of American Minorities. New York: Dodd, Mead and Company, 1975.

Sax, G. Principles of Educational Measurement and Evaluation. Belmont, California: Wadsworth Publishing Company, 1974.

Silverman, Robert J., Noa, Joslyn K., and Russell, Randall H. Oral Language Tests for Bilingual Students: An Evluation of Language Dominance and Proficiency Instruments. Portland, Oregon: Center for Bilingual Education, Northwest Regional Education Laboratory, 1977.

Smith, Richard J. and Barrett, Thomas. Teaching Reading in the Middle Grades. Rading, Massachusetts: Addison-Wesley Publishing Co., 1974.

Speevack, Yetta. The Spider Plant. New York: Antheum, 1968

Task Force Findings Specifying Remedies Available for Eliminating Past Education Practices Ruled Unlawful under Lau vs. Nichols. Washington, D.C.: Department of Health, Education, and Welfare, Summer, 1975.

Taylor, B.P. Toward a Theory of Language Acquisition. Language Learning, 1974, 24, 23-25.

Teitelbaum, Herbert and Hiller, Richard J. Bilingual Education: The Legal Mandate. Harvard Educational Review 47: 2, May, 1977, 138-170.

The Way We Go to School: The Exclusion of Children in Boston. Boston, Massachusetts: The Task Force on Children Out of School, 1970.

Thorndike, R.L. and Hagen, E.P. Measurement and Evaluation in Psychology and Education. (3rd ed.) New York: John Wiley and Sons, 1969.

U.S.Commission on Civil Rights. The Excluded Student. Washington, D.C.: United States Government Printing Office, 1972.

U.S. Commission on Civil Rights. Puerto Ricans in the Continental United States: An Uncertain Future. Washington, D.C.: United States Government Printing Office, 1976.

U.S. Commission on Civil Rights. Social Indicators of Equality for Minorities and Women. Washington, D.C.: United States Government Printing Office, 1978.

U.S. Commission on Civil Rights. The Unfinished Education: Outcome for Minority Students in Five Southwestern States. A Research Report of the Mexican American Study Series Report No. 2, Washington, D.C.: United States Government Printing Office, 1972.

Wilkinson, Andrew. The Foundations of Language. England: Oxford University Press, 1971.

Wilkinson, A.M. and Stratta, L. Listening Comprehension at Thirteen Plus. Educational Review, 1970, 22, 228-242.

Zintz, Miles V. The Reading Process: The Teacher and The Learner. (2nd ed.) Dubuque, Iowa: Wm. C. Brown Company Publishers, 1975.